

ОСОБЕННОСТИ АННОТАЦИИ «ЭЛЕКТРОННОГО ЛИТЕРАТУРНОГО КОРПУСА ДРЕВНЕГО КИТАЙСКОГО ЯЗЫКА»

Традиционно выделяют три этапа развития китайского языка: древнекитайский, среднекитайский и современный. Язык разных периодов имеет свои существенные отличия [1]. Между ними существует очевидная на сегодня разница. Современный китайский язык, являясь самодостаточной системой, обладает признаками, позволяющими говорить об его эволюции и тесной связи с древнекитайским вариантом. Особенность древнего китайского языка проявляется не только в том, что это атрибут оригинальной китайской культуры, но и в его широкой распространённости, использовании в общенациональной системе письма. Возможность изучать китайский язык в ретроспекции обусловлена наличием памятников письменности, принадлежащих разным этапам, в том числе древнекитайскому периоду истории и языка. В таком контексте есть все основания предполагать, что для всестороннего лингвистического анализа просто необходимо учитывать диахронический аспект языковой динамики.

Оцифровка древних текстов, сведение их в электронные библиотеки позволяет проводить лингвистические изыскания корпусными методами. Корпусные исследования национального языка приобретают в Китае все большую популярность, что подтверждается и активным развитием корпусных проектов, и

расширением их возможностей. С точки зрения технологии обработки языковых данных — ключевым является понятие разметки (аннотации) электронных текстовых массивов. Относительно китайского языка можно констатировать, что его уникальная грамматика предполагает дополнительный учет соответствующих особенностей в русле их адаптации к исторически сформировавшимся англо-, германоязычным и индоевропейским этимологическим традициям корпусной лингвистики.

Большинство корпусов в Китае — это, конечно, корпуса современного китайского языка. Для нашего исследования релевантным является представительный корпус, аккумулирующий древнекитайские памятники. Таким репрезентативным объектом изучения, несомненно, может являться «[汉籍电子文献资料库](#)» (Электронный литературный корпус древнего китайского языка, далее по тексту — ЭЛКДКЯ) — один из самых больших корпусов древнего китайского языка [2]. Корпус создан «[中央研究院—历史语言研究所](#)» (Центральным научно-исследовательским институтом — исследовательским институтом истории и языка) Тайваня. Собранные в этом ресурсе литературные произведения написаны на нормативном традиционном древнекитайском языке, в соответствии с канонами своего времени. Характерно, что большинство древних литературных произведений Китая сохранились именно на Тайване.

Древнекитайский язык времен Конфуция (примерно 550–480 до н. э.) в западной китаистике называется «*Archaic Chinese*», т. е. *архаический китайский язык*. На этом варианте китайского языка написаны литературные произведения того времени, в том числе: «[二十五史](#)» (Двадцать пять летописей), «[十三经](#)» (Тринадцатикнижие), более 20 миллионов нормативных документов провинции Тайвань, более 10 миллионов буддистских канонов и других древних памятников.

В 1999 г. объём ЭЛКДКЯ превысил 100 миллионов нероглифов и продолжает прирастать со скоростью около 10 миллионов нероглифов ежегодно. Такое развитие проекта позволяет идентифицировать его как динамический корпус. До 1999 г. этот корпус был открыт только для учащихся и преподавателей, но с июля 1999 г. и до сих пор он является общедоступным посредством сети Интернет.

Данный корпус снабжен разметкой. С помощью метаразметки можно найти текст по его названию, автору, содержанию, жанру, времени написания текста и т.д. Элементы собственнолингвистической разметки позволяют осуществлять поиск по лексико-семантическим параметрам. Например, можно найти отдельное слово и его синонимы. Рассмотрим сначала возможности экстралингвистической разметки ЭЛКДКЯ.

Алгоритм поиска необходимой информации в ЭЛКДКЯ содержит стандартные процедуры, позволяющие квалифицировать его как простой и удобный в использовании. Рассмотрим конкретные примеры. Выбираем необходимые параметры и «кликаем» по «кнопке» «[搜索](#)» (Поиск): откроется список текстов, согласно выбранной нами проблемной области. После чего можно перейти к функции «[阅读全文](#)» (Читать полный текст) (см. рис. 1).

Рис. 1. Библиографический обзор

Можно выбрать соответствующие рубрики *каноны, историки, мыслители, сборники поэзии и прозы*. В «окошке», которое перед нами открывается, появится список доступных книг данной категории (см. рис. 2). После «клика» по названию нужной книги система открывает полные тексты, содержащиеся в сборнике.

Рис. 2. Полный обзор библиотеки

Можно расширить критерии поиска необходимого для исследования материала. Для выбора нужной категории можно заполнять как только одно соответствующее «поле» (*заголовок, текст, комментарий*), так и несколько вариантов одновременно (см. Рисунок 3).

Рис. 3. Расширенный поиск

Рассмотрим базовые возможности собственнолингвистической разметки анализируемого корпуса (см. Рисунок 4).

【 基本检索 】

進階 專業

☐ 书名 ☒ 内文 ☒ 注释

☒ 异体字 (例如: 烟=烟)

☐ 同义词 (例如: 秦始皇=嬴政)

成书朝代限定 (不勾选表全部, 可复选)

☐ 先秦 ☐ 秦汉 ☐ 魏晋南北朝 ☐ 隋唐五代

☐ 宋辽金 ☐ 元 ☐ 明 ☐ 清 ☐ 民国

搜尋 清除

Рис. 4. Базовый поиск

В специальной «строке» набираем искомую лексему. В «окошке», которое перед нами откроется, отметим «галочкой» нужную позицию: *название, содержание, комментарий*. Если для целей исследования необходимо соответствующее представление текстовой информации, можно выбрать *традицию письма*: отмечая определённую разновидность письменности, получаем либо *традиционные иероглифы*, либо *упрощённые иероглифы*. Например, 煙-烟 — *уан* — ‘дым’, либо 學-學 — *хуе* — ‘учиться’.

Здесь же запрограммировано необходимое уточнение параметров металингвистической разметки. В частности, возможно проведение поиска с учетом соответствующих *вариантов*. Например, у одного человека могло быть несколько вариантов написания его имени: как 方苞 так и 方望溪. Также, можно выбрать соответствующую историческую эпоху — *династию*, — когда создавался или редактировался текст. Поиск может осуществляться по нескольким историческим эпохам одновременно.

Интерфейс ЭЛКДКЯ является дружелюбным и позволяет интерактивное участие пользователя в обработке запрашиваемой информации. Так, корпусный менеджер позволяет вести учет закладок и делать заметки. Можно сделать *закладку* в процессе чтения, равно как и написать *заметку*. Пользователи через учётную запись, могут открывать личные закладки / заметки в виде списка. Например, выбрав определённую *заметку*, можно увидеть фрагменты текста, к которым она относится.

Безусловно, разнообразные корпусные ресурсы существенно расширяют возможности изучения китайского языка. Корпусная лингвистика в Китае развивается впечатляющими темпами, о чем свидетельствует успешный перевод в электронный формат и аннотация памятников древнекитайского языка. Для белорусского языка, например, подобная работа даже не начиналась. Несомненно, такие возможности будут способствовать развитию как синхронических, так и

диахронических исследований китайского языка на уровне высоких стандартов современного языкознания.

ЛИТЕРАТУРА

1. Карапетянц, А.М., Тань. Учебник классического китайского языка вэньянь. Начальный курс. — М., 2001.
2. 汉籍电子文献资料库 Электронный литературный корпус древнего китайского языка [Электронный ресурс]. — URL: <http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm>. — Дата доступа: 22.02.2013.